GOODSPELL

C2B0006 MASTER

552-2060-05

apple computer inc.


SPECIAL
DELIVERY
SOFTWARE
™

GOODSPELL

C2B0006 MASTER

652-2060-0[illegible]

apple computer inc.


A P P L E
SPECIAL
DELIVERY
SOFTWARE
TM

GOODSPELL

Written by: Henry G. Baker, Ph.D.

In conjunction with Apple Computer Inc.

GOODSPELL

(We have ignored in this discussion how to encode the dictionary to also allow for efficient search).

We don't have room for 132,000 or even 90,000 characters on the Apple, so we had to do something else. Since there are only 26 letters and one separator, we could conceivably encode each character in 5 bits. This encoding would then allow the storage of our 15,000 words in only 82,500 bytes--certainly an improvement, but not nearly enough, because the Apple only has 49152 bytes of RAM, of which only about 36,000 bytes can actually be used for a program.

More complex codes would help somewhat. For example, a binary variable-length code based on character frequency would be able to encode our dictionary in close to 4 bits per character. Using this encoding, our dictionary would then require only 66,000 bytes--a savings of 22,500, but still not compact enough to fit in this small microcomputer.

We are led to consider how good the most compact encoding for our dictionary would be. It has been estimated that the information content of a character in English is only about 2 bits. Two bits per character would let us store our 15,000 words in 33,000 bytes, which would finally let us achieve our goal. However, achieving the 2 bit/character goal requires multiple character encoding, in which a single code word encodes a multi-letter sequence. A program for this encoding and decoding scheme would be complex and hopelessly slow.

The Goodspell Solution

The Goodspell dictionary fits 14,000 words in 33,690 bytes, thus achieving a density of 2.19 bits per character of dictionary--very close to our estimate of 2 bits per character for the minimum possible encoding of English.

Yet Goodspell checks files at up to 300 bytes per second--equivalent to 3000 words per minute (6 typewritten pages). Since most documents generated by word processing systems are 3 pages or less, they can be checked for correct spelling in less than one minute!

(This is to be compared with IBM's Displaywriter, which checks spelling at only 100 characters per second).

Goodspell is the solution to performance and quality in your office. Cn yu affurd to bee witot it?

Appendix A

Choice of words in main dictionary.

The 13,000 words in the main dictionary were chosen arbitrarily by Synapse to be a reasonable compromise between words which occur commonly, and "generally useful" words such as those which occur in this manual.

Thus, words such as "justifier" and "lookup" which are obviously jargon have been left out in favor of general purpose words such as "pursue", "questionable", etc.

To give you some idea as to the depth o the dictionary, Goodspell has been run on this manual itself, and the following words were not found in its main dictionary:

Goodspell Henry Bandley Cupertino misspellings eagle wallet typographical mainframe DEC MHz typewritten IBM Displaywriter Applewriter justifier autostart retry COM booting PR Dn Sn Corvus Vnnn jargon bastard filename Zipf kth logarithm lookup throughput separator cn yu affurd bee witot CF nonstandard signalled isn Jack.

(This run took 1 minute, 3 seconds.)

All the rest of the words in this manual are in the main dictionary!

apple computer inc.

GOODSPELL

C2B0006 MASTER

652-2060-00

A P P L E